

# Lyricist-Singer Entropy Affects Lyric-Lyricist Classification Performance

Mitsuki Morita, Masato Kikuchi, Tadachika Ozono

Nagoya Institute of Technology, Japan

ICAICTA 2023  
Lombok, Indonesia  
October 7-9, 2023



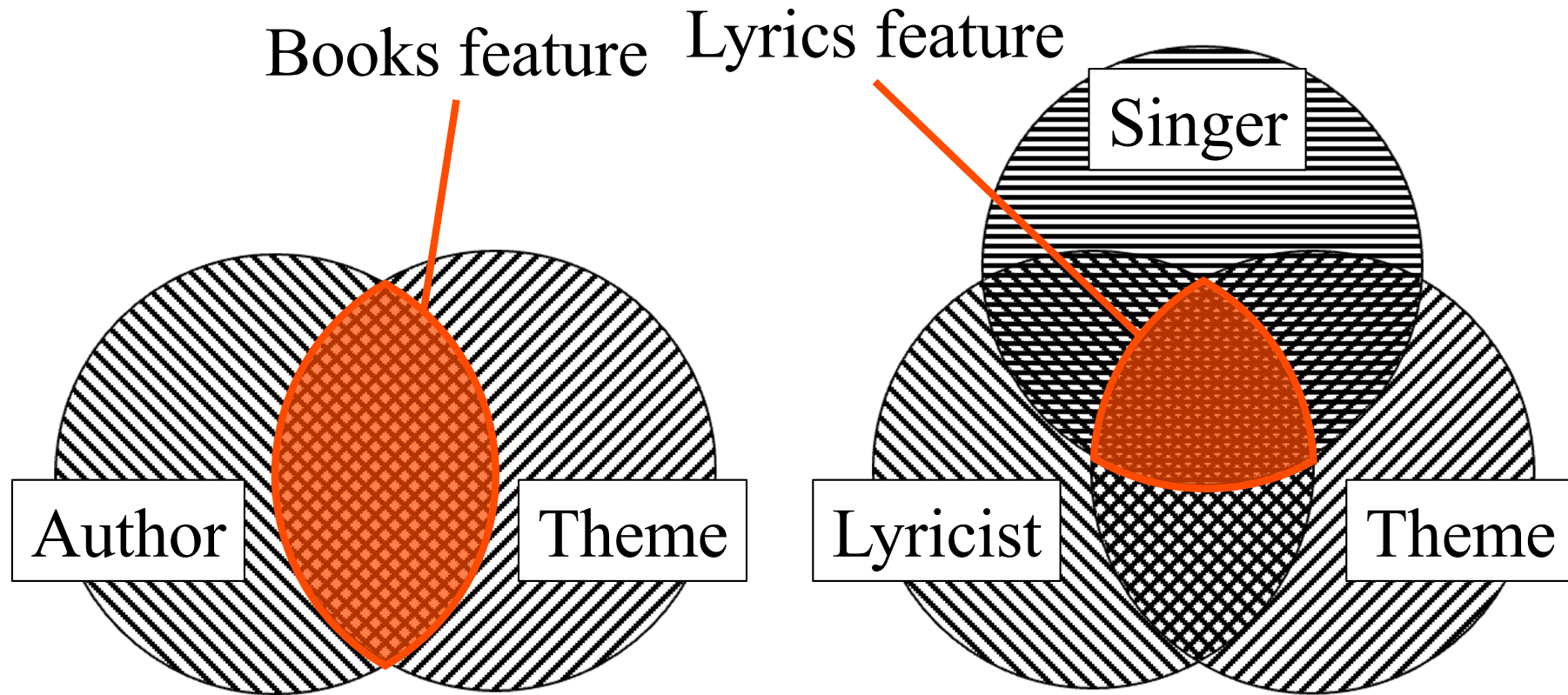
# Background

1

- The analysis of the features of the creators of literary works has been widely studied.
  - conducted for many purposes.
- We focused on the lyricists who are the creators of the lyrics.
  - However, lyricists have received less attention than singers and lyrics.

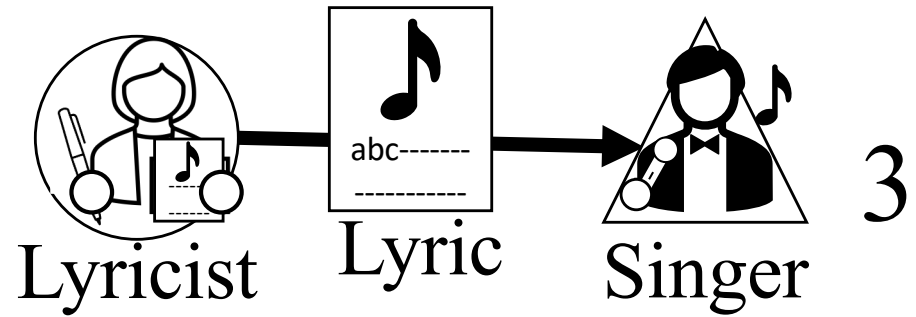
# Lyric Specificity

2



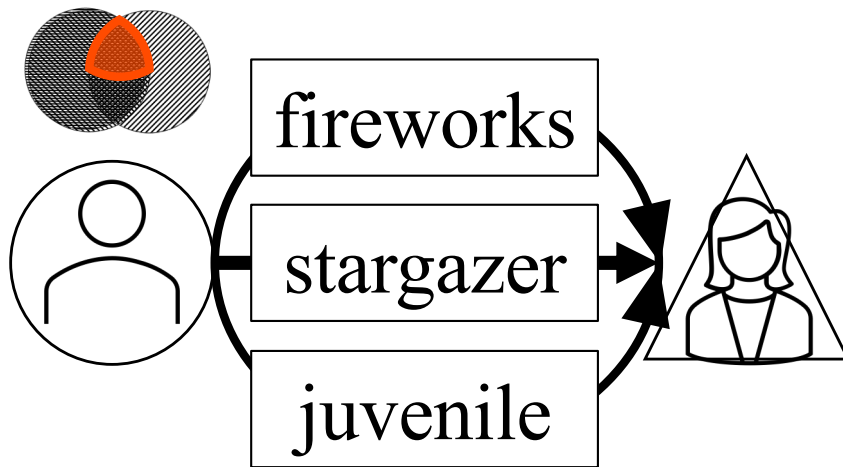
Which is difficult to capture the features?

# Lyricist and Singer

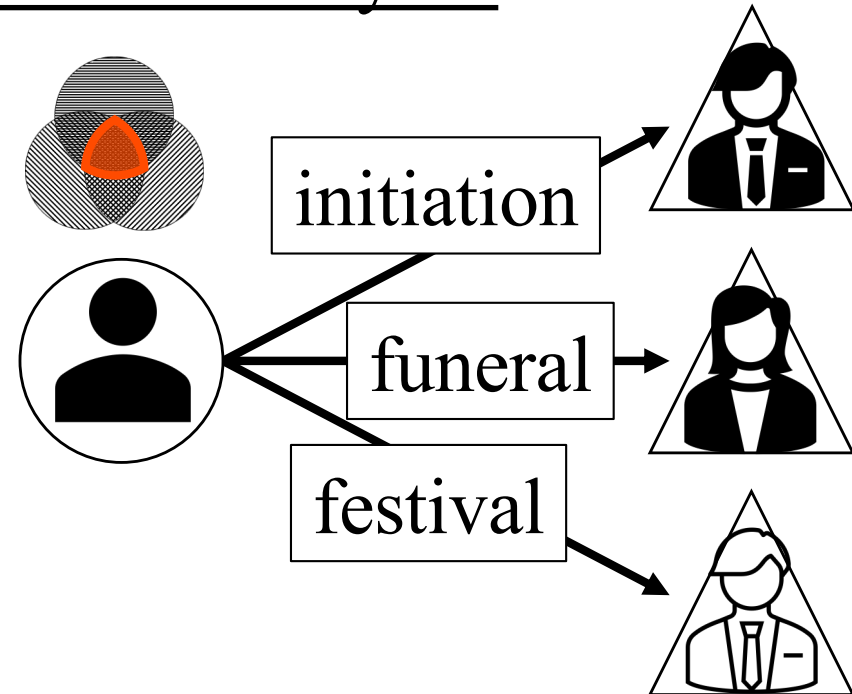


Which lyricist is easier to analyze?

lyricist = singer



The left lyricist wrote for a single singer



The right lyricist wrote for multiple singers

# Motivation and Procedure

4

- We hypothesized that lyricists with fewer associated singers would be more likely to capture the features.
- We investigated relationship between two elements.
  - The variety of singers associated with the lyricist.
    - : Quantify as **Lyricist-Singer entropy**.
  - The ease of capturing the features of the lyricist.
    - : Quantify as **Lyric-Lyricist classification performance**.

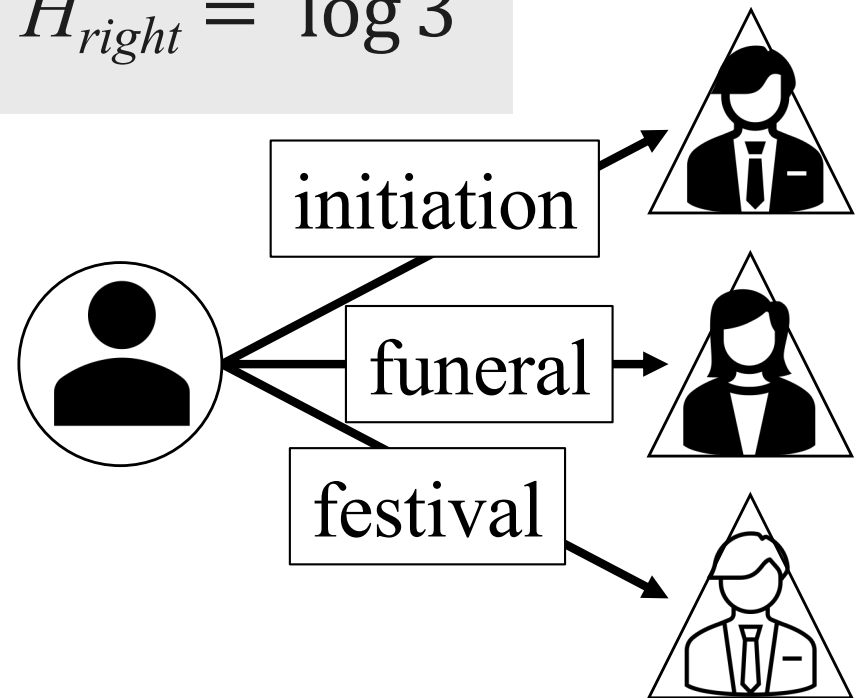
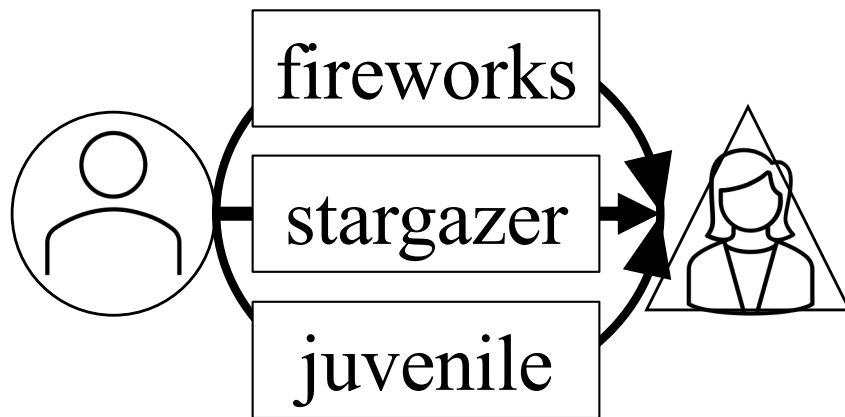
# Lyricist-Singer Entropy

5

Lyricist-Singer Entropy:  
(LS Entropy)

$$H_i = - \sum_j \frac{|X_i \cap X_j|}{|X_i|} \log \frac{|X_i \cap X_j|}{|X_i|}$$

$$H_{left} = 0 < H_{right} = \log 3$$



# Classification task

6

- We conducted Lyric-Lyricist classification.
  - input: lyric text
  - output: lyricist who seems to write the input lyrics
- Performance is calculated from the number of correct and incorrect answers.
  - Higher classification performance shows that it is easy for us to capture the characteristics of lyricists.
  - Lower classification performance shows that it is hard for us to capture the characteristics of lyricists.

# Experimental Procedure

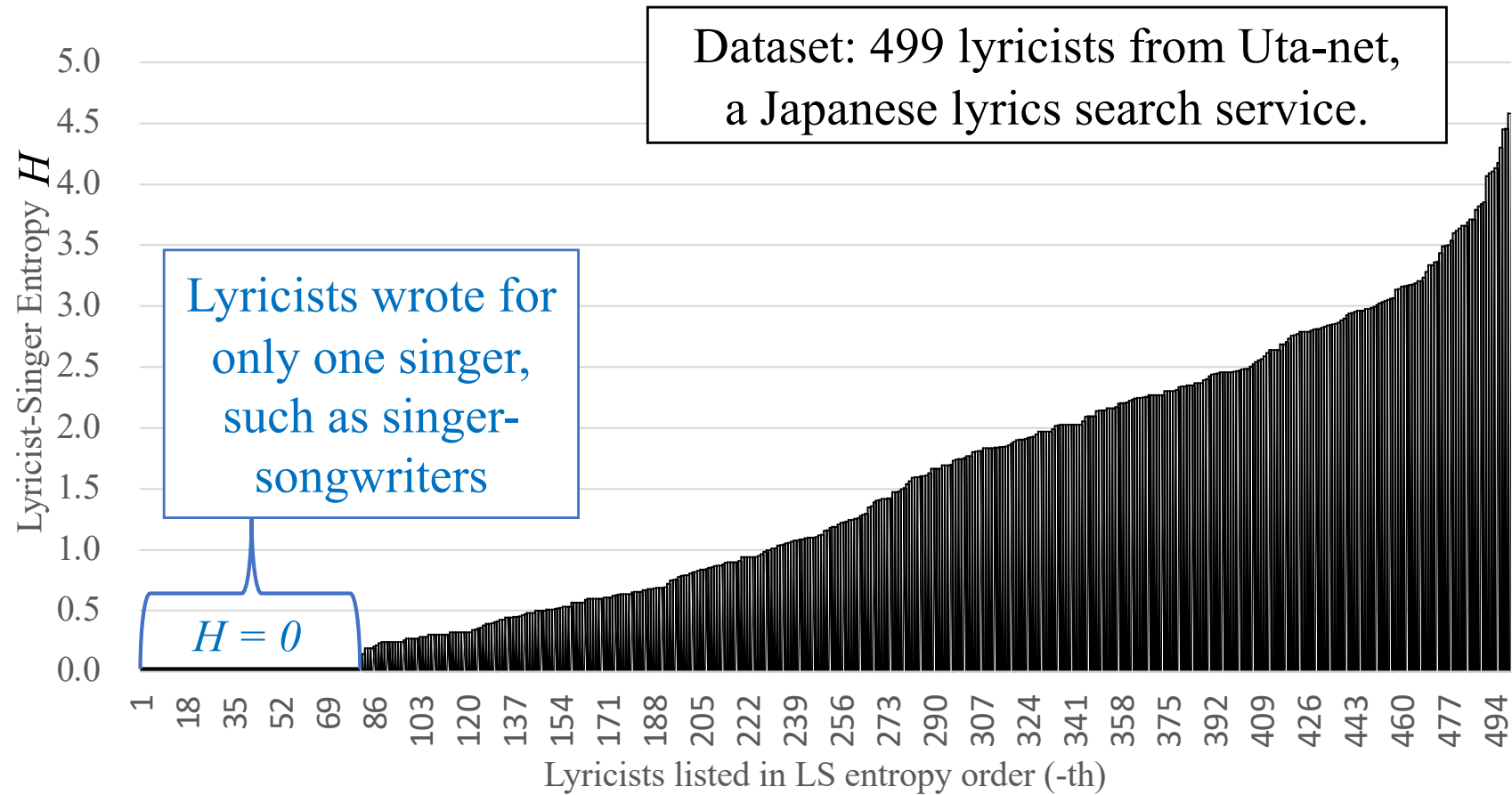
7

- Investigated the relationship between LS entropy and LL classification performance.
  - By grouping the lyricists by LS entropy and comparing the LL classification performance of each group.
- The experiment was conducted using as follows.
  1. Data Collection
  2. Grouping Lyricist (using two grouping methods)
  3. Dataset Constructing (using two sampling methods)



# Lyricist Distribution

8



# Classification Method

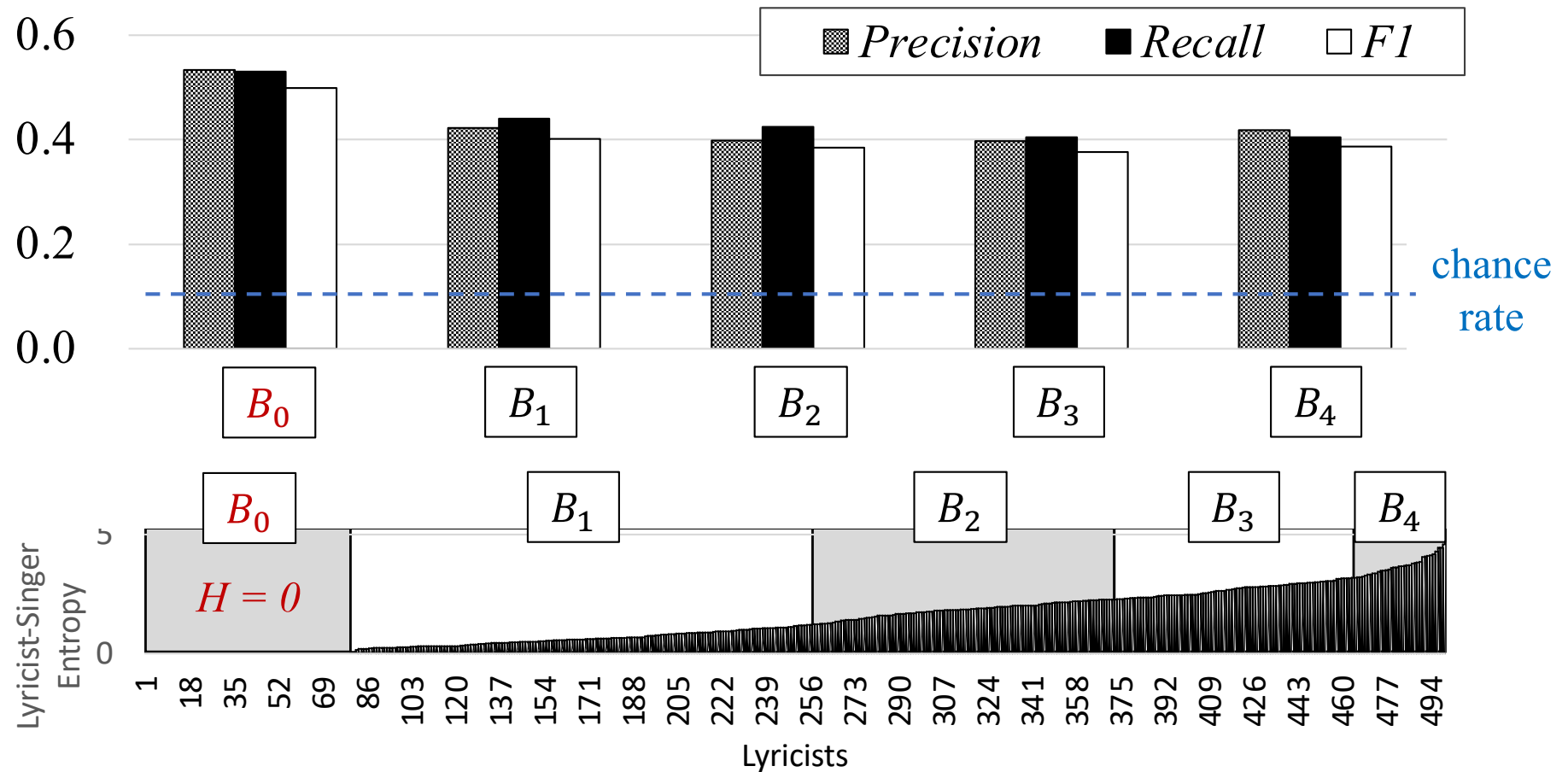
9

- Our classifier fine-tuned a BERT model pre-trained on Japanese texts.
  - input: One-hot vector of **lyrics**, first 512 tokens.
  - output: 10-dimensional vector of **lyricist**'s probabilities.
- The procedure for constructing datasets is as follows:
  1. Select lyricists (using two sampling methods)
  2. Select songs
  3. Separate into training dataset, evaluation dataset, and test dataset

# Experimental Result

10

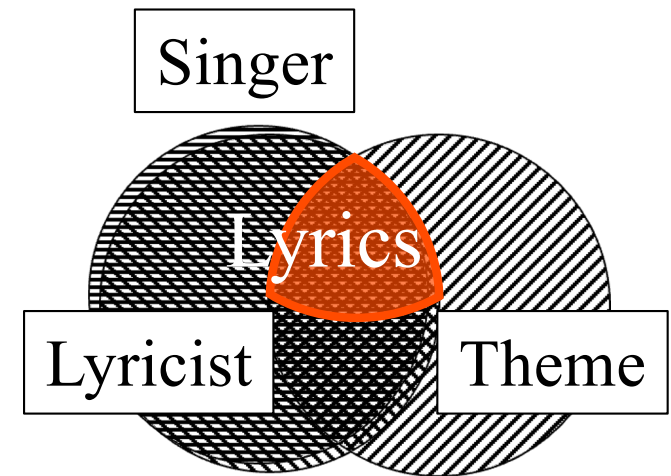
Lower LS entropy ( $H$ ) , higher classification performance



# Discussion

11

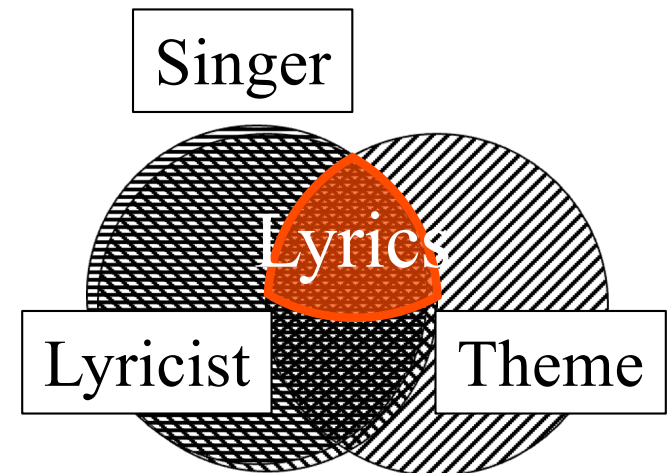
- Our experimental results suggest we can more easily capture the characteristics of lyricists who wrote for fewer singers because they are more classifiable.
  - The lower  $H$  group had higher classification performance.
  - Specifically, lyricists who wrote lyrics for only one specific singer,  $H = 0$ , had the best classification performance.



# Conclusion

12

- Our experimental results suggest  $H = 0$  lyricists, such as singer-songwriters, are promising for capturing their characteristics.
- We investigated the relationship between singer variety (LS entropy,  $H$ ) and the ease of capturing features (LL classification performance)
- The zero LS entropy,  $H = 0$ , group had the highest F1 score, 0.499.



# References

13

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, volume 1, pages 4171–4186, 2019.
- [2] Wanwan Zheng and Mingzhe Jin. A review on authorship attribution in text mining. *WIREs Computational Statistics*, 15(2), 2023.
- [3] Mae'i Fabien, Esau Villatoro-Tello, Petr Motlicek, and Shantipriya Parida. BertAA : BERT fine-tuning for authorship attribution. In *Proceedings of the 17th International Conference on Natural Language Processing (ICON)*, pages 127–137, 2020.
- [4] Ankita Dhar, Himadri Mukherjee, Shibaprasad Sen, Md Obaidullah Sk, Amitabha Biswas, Teresa Gonçalves, and Kaushik Roy. Author identification from literary articles with visual features: A case study with bangla documents. *Future Internet*, 14(10):1–20, 2022.
- [5] Yunita Sari, Mark Stevenson, and Andreas Vlachos. Topic or style? exploring the most useful features for authorship attribution. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 343–353, 2018.
- [6] Anastasia Fedotova, Aleksandr Romanov, Anna Kurtukova, and Alexander Shelupanov. Digital authorship attribution in russian-language fanfiction and classical literature. *Algorithms*, 16(1):1–32, 2023.
- [7] Silvia Corbara, Alejandro Moreo, and Fabrizio Sebastiani. Syllabic quantity patterns as rhythmic features for latin authorship attribution. *Journal of the Association for Information Science and Technology*, 74(1):128–141, 2023.
- [8] Makarand Velankar, Rachita Kotian, and Parag Kulkarni. Contextual mood analysis with knowledge graph representation for hindi song lyrics in devanagari script. *CoRR*, abs/2108.06947, 2021.
- [9] Silvia Corbara and Alessio Molinari. Reading songs: A computational analysis of popular songs lyrics. In *Proceedings of the 11th Italian Information Retrieval Workshop 2021*.
- [10] Daichi Haraguchi, Shota Sakaguchi, Jun Kato, Masataka Goto, and Seiichi Uchida. Fonts that fit the music: A multimodal design trend analysis of lyric videos. *IEEE Access*, 10:65414–65425, 2022.
- [11] Tunç Yılmaz and Tatjana Scheffler. Song authorship attribution: a lyrics and rhyme based approach. *International Journal of Digital Humanities*, 5(1):29–44, 2023.